

ICAE

Instituto Complutense de Análisis Económico

UNIVERSIDAD COMPLUTENSE

FACULTAD DE ECONOMICAS

Campus de Somosaguas

28223 MADRID

Teléfono 394 26 11 - FAX 294 26 13



UU
49
(9802)

Documento de trabajo

**Un algoritmo rápido para evaluar la
función de verosimilitud exacta de
modelos VARMAX periódicos**

José Casals

Sonia Sotoca

Miguel Jerez

No. 9802

Mayo 1998

ICAE

Instituto Complutense de Análisis Económico

UNIVERSIDAD COMPLUTENSE

**UN ALGORITMO RÁPIDO PARA EVALUAR LA FUNCIÓN DE
VEROSIMILITUD EXACTA DE MODELOS VARMAX PERIÓDICOS**

José Casals
Sonia Sotoca
Miguel Jerez

Universidad Complutense de Madrid
Campus de Somosaguas
28223 Madrid

ABSTRACT

In this work we derive a fast algorithm to compute the exact likelihood function of periodic VARMAX processes. Its computational efficiency is achieved by combining a minimal dimension state-space formulation, in steady-state innovations form, and a procedure for computing the exact likelihood function which takes advantage of the properties of this representation. The algorithm can be applied to stationary and non-stationary models, allows for deterministic and/or stochastic exogenous variables and makes easy the computation of the exact second-order moments of the estimates. On the other hand, our approach includes representations not considered by the literature, like multivariate periodic processes, and allows for nonhomogeneous dynamic structures and different number of observations in each season. Besides, it can be applied to any model with deterministic parameter variation. Some results with simulated data illustrate the good behaviour of the algorithm.

RESUMEN

En este trabajo se deriva un algoritmo rápido para evaluar la función de verosimilitud exacta de procesos VARMAX periódicos. Su eficiencia computacional se consigue combinando una formulación de dimensión mínima en espacio de los estados, en forma *steady-state innovations* y un procedimiento para evaluar la función de verosimilitud exacta que aprovecha las propiedades de esta representación. El algoritmo es aplicable a modelos estacionarios y no estacionarios, con variables exógenas estocásticas y/o deterministas y facilita el cálculo de los segundos momentos exactos de las estimaciones. Por otra parte, la representación utilizada permite tratar casos no considerados en la literatura, como procesos periódicos multivariantes, y admite estructuras dinámicas no homogéneas y muestras con distinto número de observaciones en cada estación. Asimismo, es inmediatamente aplicable a cualquier caso de variación paramétrica determinista. Algunas pruebas con datos simulados ponen de manifiesto el buen funcionamiento del algoritmo.

Key words: periodic VARMAX, exact maximum likelihood, Kalman filter, steady-state innovations models, state-space models.

JEL classification codes: C32; C40.

Mailing address: Departamento de Economía Cuantitativa. Universidad Complutense. Somosaguas. 28223 Madrid. Spain. Fax: 394 26 13. E-mail: eccua01@sis.ucm.es.

h.c. : 53-314327-5

h.e. : 534058708

1. Introducción.

Como es bien sabido, la estimación por máxima verosimilitud exacta de modelos econométricos de series temporales puede abordarse aplicando el filtro de Kalman a su representación equivalente en Espacio de los Estados (EE) (Gardner *et al.*, 1980; Terceiro, 1990). Esta aproximación tiene distintas ventajas, derivadas de su capacidad para resolver problemas de estimación en situaciones estándar y no estándar. En este trabajo se aplica este enfoque a la estimación por máxima verosimilitud exacta de modelos periódicos.

Los modelos con variación paramétrica estacional son útiles para modelizar y predecir series estacionales (Noakes *et al.*, 1985). Sin embargo, su estimación es más compleja y costosa que la de modelos ARMA estacionales, sobre todo por su elevado número de parámetros.

Pagano (1978) aborda la estimación de modelos periódicos autorregresivos (PAR) mediante el método de los momentos. En la misma línea, Salas *et al.* (1982) utilizan las ecuaciones de Yule-Walker estacionales para estimar modelos PARMA.

Como indica Vecchia (1985), una manera de estimar un proceso PARMA consiste en obtener su representación equivalente en forma VARMA y estimarlo por máxima verosimilitud. El inconveniente de este procedimiento es su elevado coste computacional, tanto en términos de almacenamiento como de tiempo de cálculo. Por ello, este autor sugiere un método computacionalmente eficiente que, sin embargo, no proporciona estimaciones exactas cuando el modelo incluye factores autorregresivos. Li y Hui (1988) resuelven esta deficiencia derivando un algoritmo de máxima verosimilitud exacta para modelos PARMA estacionarios, que extiende los resultados de McLeod (1975) al caso periódico.

Otros trabajos recientes se centran en el desarrollo de algoritmos rápidos para la estimación de modelos PARMA utilizando un enfoque recursivo. En esta línea, Adams y Goodwin (1995) proponen un método que proporciona estimaciones consistentes aunque no eficientes. Por su parte, Boshnakov (1996) adapta el procedimiento recursivo de Hannan y Rissanen (1982) y Hannan y Kavalieris (1984) al caso de modelos PARMA estacionarios, permitiendo estimar simultáneamente el orden del proceso y sus parámetros.

Los métodos mencionados en los párrafos anteriores: a) sólo contemplan el caso univariante y estacionario, b) no consideran la posible existencia de variables exógenas en el modelo, c) suponen el mismo número de observaciones por estación y d) cuando no todas las estaciones presentan la misma estructura dinámica incurren en ineficiencia, ya que necesitan aumentar artificialmente la dinámica del sistema.

En este trabajo se propone una formulación generalizada - proceso VARMAX periódico o PVARMAX - que evita la mayoría de estas limitaciones. Asimismo, se desarrolla un algoritmo computacionalmente eficiente para estimarlo por máxima verosimilitud exacta. La eficiencia se consigue combinando una representación de dimensión mínima en EE en forma *steady-state innovations*, y un algoritmo para evaluar la función de verosimilitud exacta que aprovecha las propiedades de convergencia del filtro de Kalman en este tipo de modelos.

El uso de una representación en EE tiene ventajas adicionales, ya que facilita la aplicación de resultados existentes para tratar modelos estacionarios o no estacionarios, con variables explicativas estocásticas y/o deterministas (Casals y Sotoca, 1997), detectar automáticamente raíces unitarias y calcular los segundos momentos exactos de las estimaciones (Terceiro, 1990 y Apéndice).

La estructura del artículo es la siguiente. En la Sección 2 se define el proceso generalizado PVARMAX. En la Sección 3 se deriva su formulación de dimensión mínima en EE y se establece un conjunto de condiciones necesarias y suficientes de estacionariedad e invertibilidad. La Sección 4 presenta un algoritmo eficiente para evaluar la función de verosimilitud exacta de un modelo PVARMAX en EE, siguiendo a De Jong (1988). En la Sección 5 se resume la estructura del algoritmo y sus principales propiedades. La validez del procedimiento se comprueba en la Sección 6, simulando modelos análogos a los utilizados por Li y Hui (1988) y Vecchia (1985), en el caso univariante, y estructuras comparables en el caso multivariante. Finalmente, en la Sección 7 se resumen las conclusiones del trabajo y en el Apéndice se deriva la matriz de información exacta de un proceso PVARMAX, así como la forma de calcularla eficientemente.

2. Definición del proceso PVARMAX.

Sean $\{z_t\}$ un vector $(m \times 1)$ de series temporales estacionales y $\{u_t\}$ un vector $(r \times 1)$ de variables exógenas, deterministas y/o estocásticas de tamaño N . Se dice que $\{z_t\}$ sigue un proceso PVARMAX $_s(p, q, g)$, $\forall i = 1, 2, \dots, S$, donde S es el número de períodos muestrales que forman una estación completa, si:

$$F_t(B)z_t = G_t(B)u_t + \Theta_t(B)a_t \quad (1)$$

siendo

$$F_t(B) = I + F_{t,1}B + F_{t,2}B^2 + \dots + F_{t,p}B^p \quad (2)$$

$$\Theta_t(B) = I + \Theta_{t,1}B + \Theta_{t,2}B^2 + \dots + \Theta_{t,q}B^q \quad (3)$$

$$G_t(B) = G_{t,0} + G_{t,1}B + G_{t,2}B^2 + \dots + G_{t,g}B^g \quad (4)$$

y, para una muestra de tamaño N , los coeficientes cambian de acuerdo con una ley de variación determinista de período S :

$$F_t(B) = F_{t,s}(B); G_t(B) = G_{t,s}(B); \Theta_t(B) = \Theta_{t,s}(B) \quad (5)$$

$$E(a_t a_t^T) = \Sigma_t; \Sigma_t = \Sigma_{t,s} \quad t = 1, 2, \dots, N \quad (6)$$

A los parámetros de $F_t(B)$ se les denomina parámetros AR estacionales, los de $\Theta_t(B)$ son los parámetros MA estacionales y los de Σ_t son las varianzas estacionales del error.

Por tanto, un proceso PVARMAX puede interpretarse como un modelo VARMAX con matrices de coeficientes que cambian en el tiempo de forma periódica (ver Lütkepohl, 1993).

El modelo (1)-(6) incluye como casos particulares las formulaciones periódicas convencionales. Por ejemplo, un proceso PARMA $_s(p, q)$, es un caso particular de (1)-(6) con $m=1$ y $g=0$. Es decir:

$$f_t(B)z_t = \theta_t(B)a_t \quad (7)$$

$$f_t(B) = 1 + f_{t,1}B + f_{t,2}B^2 + \dots + f_{t,p}B^p \quad (8)$$

$$\theta_t(B) = 1 + \theta_{t,1}B + \theta_{t,2}B^2 + \dots + \theta_{t,q}B^q \quad (9)$$

$$f_t(B) = f_{t,s}(B); \theta_t(B) = \theta_{t,s}(B); E(a_t a_t^T) = \sigma_t^2, \sigma_t^2 = \sigma_{t,s}^2, t = 1, 2, \dots, N \quad (10)$$

3. Formulación en EE de un proceso PVARMAX.

Si se abordara directamente la estimación de un modelo en la forma (1)-(4) y no todas las estaciones presentaran la misma estructura dinámica, el procedimiento incurriría en la misma ineficiencia computacional que los procedimientos convencionales. En esta sección se discutirá una norma para representar el proceso en forma de EE que garantiza que la dinámica es la mínima imprescindible.

Las ecuaciones (1)-(4) pueden representarse en forma de EE como:

$$x_{t+1} = \Phi_t x_t + \Gamma_t u_t + E_t a_t \quad (11)$$

$$z_t = H_t x_t + D_t u_t + a_t \quad (12)$$

donde el vector de perturbaciones a_t tiene una matriz de covarianzas Q_t y $\Phi_t, \Gamma_t, E_t, H_t$ y D_t son matrices de tamaño $(n_{t+1} \times n_t)$, $(n_{t+1} \times r)$, $(n_{t+1} \times m)$, $(m \times n_t)$ y $(m \times r)$, respectivamente, que cambian en el tiempo de acuerdo con una ley de variación periódica $\Phi_t = \Phi_{t,s}, \Gamma_t = \Gamma_{t,s}, E_t = E_{t,s}, H_t = H_{t,s}, D_t = D_{t,s}$ y $Q_t = Q_{t,s}, t = 1, 2, \dots, N$.

Las expresiones (11)-(12) caracterizan un modelo en EE en forma *steady-state innovations*. Como se verá en la sección 4, cuando se aplica el filtro de Kalman para evaluar la función de verosimilitud exacta de este tipo de modelos, presenta propiedades de convergencia que pueden aprovecharse para aumentar sustancialmente su eficiencia computacional.

Para determinar la relación entre el modelo (1)-(4) y las matrices de (11)-(12) conviene distinguir entre el caso en que los s modelos tienen una misma estructura dinámica, aunque sus parámetros puedan ser distintos entre sí, y aquél otro en que la estructura dinámica correspondiente a distintas estaciones es diferente.

3.1. Formulación cuando todos los modelos tienen la misma estructura dinámica.

Cuando los modelos de todas las estaciones tienen la misma estructura dinámica, se cumple $k_i = k, \forall i = 1, 2, \dots, s$, siendo $k_i = \max\{p_i, q_i, g_i\}$. En este caso, la norma de conversión del modelo (1)-(4) a la forma (11)-(12) es:

$$\Phi_t = \begin{bmatrix} -F_{t,1} & I & 0 & \dots & 0 \\ -F_{t+1,2} & 0 & I & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -F_{t+k-2,k-1} & 0 & 0 & \dots & I \\ -F_{t+k-1,k} & 0 & 0 & \dots & 0 \end{bmatrix}; \quad \Gamma_t = \begin{bmatrix} G_{t,1} - F_{t,1} G_{t,0} \\ G_{t+1,2} - F_{t+1,2} G_{t,0} \\ \vdots \\ G_{t+k-2,k-1} - F_{t+k-2,k-1} G_{t,0} \\ G_{t+k-1,k} - F_{t+k-1,k} G_{t,0} \end{bmatrix} \quad (13)$$

$$E_t = \begin{bmatrix} \Theta_{t,1} - F_{t,1} \\ \Theta_{t+1,2} - F_{t+1,2} \\ \vdots \\ \Theta_{t+k-2,k-1} - F_{t+k-2,k-1} \\ \Theta_{t+k-1,k} - F_{t+k-1,k} \end{bmatrix}; \quad H_t = [I \ 0 \ \dots \ 0]; \quad D_t = G_{t,0}; \quad Q_t = R_t = S_t = E(a_t a_t^T) \quad (14)$$

donde $F_{t+j,j+1} = F_{t+j+1,j+1}$; $G_{t+j,j+1} = G_{t+j+1,j+1}$; $\Theta_{t+j,j+1} = \Theta_{t+j+1,j+1}$, $\forall j=0,1,\dots,k-1$.

Ejemplo. Si $s = 3$ y la estructura dinámica en cada estación está caracterizada por un proceso AR(1). La representación convencional del modelo PAR₃(1) es:

$$\begin{aligned} z_{t+1} &= f_{1,1} z_t + a_{t+1} \\ z_{t+2} &= f_{2,1} z_{t+1} + a_{t+2} \\ z_{t+3} &= f_{3,1} z_{t+2} + a_{t+3} \quad t=0,3,6,\dots \end{aligned} \quad (15)$$

y la representación de cada uno de estos modelos en EE es, respectivamente:

$$\begin{aligned} z_{t+1} &= x_{t+1} + a_{t+1} \\ x_{t+2} &= f_{2,1} x_{t+1} + f_{2,1} a_{t+1} \end{aligned} \quad (16)$$

$$\begin{aligned} z_{t+2} &= x_{t+2} + a_{t+2} \\ x_{t+3} &= f_{3,1} x_{t+2} + f_{3,1} a_{t+2} \end{aligned} \quad (17)$$

$$\begin{aligned} z_{t+3} &= x_{t+3} + a_{t+3} \\ x_{t+4} &= f_{1,1} x_{t+3} + f_{1,1} a_{t+3} \end{aligned} \quad (18)$$

3.2. Formulación cuando no todos los modelos tienen la misma estructura dinámica.

Cuando los s modelos no tienen la misma estructura dinámica $k_1 \neq k_2 \neq \dots \neq k_s$, la norma de conversión (13)-(14) es válida, sin más que ampliar artificialmente la dinámica para que todas las estaciones tengan la misma estructura. El problema es que la aplicación de (13)-(14) da lugar a una representación en EE de dimensión no mínima.

En la representación en EE de dimensión mínima el número de filas de Φ_t es $n_{t+1} = m \sum_{j=1}^s (k_j^+ + k_j^-)$, donde:

- 1) k_j^+ es la parte entera de $\frac{k_j}{s}$
- 2) $k_j^- = \begin{cases} 1 & \text{si } k_j - k_j^+ s \geq j - t \text{ cuando } j > t \\ 1 & \text{si } k_j - k_j^+ s \geq j + s - t \text{ cuando } j \leq t \\ 0 & \text{en caso contrario} \end{cases} \quad (19)$

siendo fácil comprobar que $n_{t+1} = n_{t+1+s}$, $\forall t = 1, 2, \dots, N$. La norma de conversión para obtener una formulación de dimensión mínima consiste en eliminar los estados redundantes, es decir, aquéllos que no estén afectados por ningún input observable (u_t) o no observable (a_t). La forma de construir las matrices Φ , Γ y E es la siguiente. Sea $k = \max\{k_i\}$, $\forall i = 1, 2, \dots, s$. Entonces, si $k_{i,j} \geq j$ para $j = 1, 2, \dots, k$:

- 1) La matriz Γ se construye añadiendo los bloques de filas $[G_{i,jj} - F_{i,jj} G_{i,0}]$.
- 2) La matriz E se construye añadiendo los bloques de filas $[\Theta_{i,jj} - F_{i,jj}]$.
- 3) La matriz Φ se construye añadiendo los bloques de filas $[-F_{i,jj} \ 0 \ \dots \ 0]$.

Si $k_{i,j} < j$ para $j = 1, 2, \dots, k$, se añaden a la matriz Φ las columnas $\begin{bmatrix} 0 \\ 0 \\ \vdots \\ I \end{bmatrix}$.

Ejemplo: Si $s = 3$, $k_1 = 6$, $k_2 = 1$ y $k_3 = 1$. Siguiendo la norma de conversión (13)-(14), la representación en EE de dimensión no mínima es:

$$\Phi_1 = \begin{bmatrix} -F_{21} & I & 0 & 0 & 0 & 0 \\ -F_{32} & 0 & I & 0 & 0 & 0 \\ -F_{13} & 0 & 0 & I & 0 & 0 \\ -F_{24} & 0 & 0 & 0 & I & 0 \\ -F_{35} & 0 & 0 & 0 & 0 & I \\ -F_{16} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \Phi_2 = \begin{bmatrix} -F_{31} & I & 0 & 0 & 0 & 0 \\ -F_{12} & 0 & I & 0 & 0 & 0 \\ -F_{23} & 0 & 0 & I & 0 & 0 \\ -F_{34} & 0 & 0 & 0 & I & 0 \\ -F_{15} & 0 & 0 & 0 & 0 & I \\ -F_{26} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \Phi_3 = \begin{bmatrix} -F_{11} & I & 0 & 0 & 0 & 0 \\ -F_{22} & 0 & I & 0 & 0 & 0 \\ -F_{33} & 0 & 0 & I & 0 & 0 \\ -F_{14} & 0 & 0 & 0 & I & 0 \\ -F_{25} & 0 & 0 & 0 & 0 & I \\ -F_{36} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (20)$$

y la aplicación de la regla (19) elimina los estados redundantes, proporcionando las siguientes matrices con dimensión mínima:

$$\Phi_1 = \begin{bmatrix} -F_{21} & 0 \\ -F_{13} & I \\ -F_{16} & 0 \end{bmatrix}; \Phi_2 = \begin{bmatrix} -F_{31} & 0 & 0 \\ -F_{12} & I & 0 \\ -F_{15} & 0 & I \end{bmatrix}; \Phi_3 = \begin{bmatrix} -F_{11} & I & 0 \\ -F_{14} & 0 & I \end{bmatrix} \quad (21)$$

La aplicación del método descrito para eliminar estados redundantes hace que en la matriz Φ_1 dada en (21) desaparezcan las filas que contienen F_{32} , F_{24} y F_{35} y las columnas 2, 3, 5 y 6. En Φ_2 se eliminan las filas asociadas a F_{23} , F_{34} y F_{26} y las columnas 2, 4 y 5. Finalmente, en Φ_3 desaparecen las filas correspondientes a F_{22} , F_{33} , F_{25} y F_{36} y las columnas 3, 4 y 6.

3.3. Condiciones de estacionariedad e invertibilidad.

Las condiciones de estacionariedad e invertibilidad de un proceso periódico suelen caracterizarse a partir de la representación VARMA equivalente de parámetros fijos (ver, por ejemplo, Franses y Paap, 1994; Osborn, 1991; Boswijk y Franses, 1995). Procediendo de la misma forma, para caracterizar las condiciones de estacionariedad de un proceso PVARMAX se escribe la ecuación de estado (11) en la forma equivalente:

$$x_{t+s} = (\Phi_{t+s-1} \Phi_{t+s-2} \dots \Phi_t) x_t + \sum_{i=1}^s [(\prod_{j=1}^{s-i} \Phi_{t+j-1}) (\Gamma_{t+i-1} u_{t+i-1} + E_{t+i-1} a_{t+i-1})] \quad (22)$$

cuyas matrices no cambian cada s períodos debido a la ley de variación: $\Phi_t = \Phi_{t+s}$, $\Gamma_t = \Gamma_{t+s}$, $E_t = E_{t+s}$, $t=1, 2, \dots, N$. Una vez escrito el modelo en la forma (22)-(12), se dice que es ciclo-estacionario si todos los autovalores de la matriz $\Phi^* = \prod_{j=1}^s \Phi_{t+j-1}$ son, en módulo, menores que la unidad (ver Anderson y Moore, 1979).

Para caracterizar las condiciones de invertibilidad del modelo PVARMAX, es conveniente utilizar su representación equivalente VARX(∞). En concreto, el modelo (11)-(12) puede escribirse como:

$$x_{t+1} = \bar{\Phi}^* x_{t+1-s} + \bar{\Gamma}^* \begin{bmatrix} u_t \\ u_{t-1} \\ \vdots \\ u_{t+1-s} \end{bmatrix} + \bar{E}^* \begin{bmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t+1-s} \end{bmatrix} \quad (23)$$

$$z_{t+1} = H_{t+1} \bar{\Phi}^* x_{t+1-s} + H_{t+1} \bar{\Gamma}^* \begin{bmatrix} u_t \\ u_{t-1} \\ \vdots \\ u_{t+1-s} \end{bmatrix} + H_{t+1} \bar{E}^* \begin{bmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t+1-s} \end{bmatrix} + D_t u_t + a_{t+1} \quad (24)$$

en donde:

$$\begin{aligned} \bar{\Phi}^* &= (\Phi_t - E_t H_t) (\Phi_{t-1} - E_{t-1} H_{t-1}) \dots (\Phi_{t-s+1} - E_{t-s+1} H_{t-s+1}) \\ \bar{E}^* &= [E_t, (\Phi_t - E_t H_t) E_{t-1}, (\Phi_t - E_t H_t) (\Phi_{t-1} - E_{t-1} H_{t-1}) E_{t-2}, \dots, \\ &\quad \prod_{i=0}^{s-2} (\Phi_{t-i} - E_{t-i} H_{t-i}) E_{t+1-s}] \\ \bar{\Gamma}^* &= [\bar{\Gamma}_t, (\Phi_t - E_t H_t) \bar{\Gamma}_{t-1}, (\Phi_t - E_t H_t) (\Phi_{t-1} - E_{t-1} H_{t-1}) \bar{\Gamma}_{t-2}, \dots, \\ &\quad \prod_{i=0}^{s-2} (\Phi_{t-i} - E_{t-i} H_{t-i}) \bar{\Gamma}_{t+1-s}], \text{ con } \bar{\Gamma}_t = \Gamma_t - E_t D_t \end{aligned}$$

Puesto que la ecuación (24) puede interpretarse como un proceso VARX(∞) sustituyendo recursivamente x_{t+1-s} por su expresión (23), resulta inmediato ver que para que los valores pasados de $\{z_t\}$ tengan un efecto decreciente sobre el valor actual (concepto habitual de invertibilidad) debe cumplirse que los autovalores de $\bar{\Phi}^*$ sean menores en módulo que la unidad.

4. La función de verosimilitud exacta de un modelo PVARMAX.

Supongamos que a) la matriz de observaciones $Z = [z_1, z_2, \dots, z_N]$ ha sido generada por el modelo (11)-(12), b) las perturbaciones w_t y v_t y el estado inicial x_1 son independientes y siguen una distribución normal y c) x_t tiene una esperanza desconocida, \bar{x}_1 , y una matriz de covarianzas, P_1 , también desconocida.

En estas condiciones, la función soporte de verosimilitud puede expresarse como (ver De Jong, 1988):

$$l(Z/U, \theta) = \log |P_1| + \bar{x}_1^T P_1^{-1} \bar{x}_1 + \sum_{t=1}^N \log |B_t| + \sum_{t=1}^N \bar{z}_t^T B_t^{-1} \bar{z}_t + \log |P_1^{-1} + W_N| - [P_1^{-1} \bar{x}_1 + w_N]^T [P_1^{-1} + W_N]^{-1} [P_1^{-1} \bar{x}_1 + w_N] \quad (25)$$

La evaluación de (25) puede realizarse en tres fases claramente diferenciadas: a) cálculo de \bar{z}_t y B_t , b) cálculo de w_N y W_N y c) determinación de las condiciones iniciales \bar{x}_1 y P_1 .

Los valores de \bar{z}_t y B_t pueden obtenerse propagando un filtro de Kalman a partir de un estado inicial y una incertidumbre asociada nulos, proceso que denotaremos a partir de ahora FK(0,0). Esto da lugar al siguiente proceso recursivo:

$$\hat{x}_{t+1|t} = \Phi_t \hat{x}_{t|t-1} + \Gamma_t u_t + K_t \bar{z}_t \quad (26)$$

$$\bar{z}_t = z_t - H_t \hat{x}_{t|t-1} - D_t u_t \quad (27)$$

$$B_t = Q_t \quad (28)$$

$$K_t = E_t \quad (29)$$

$$P_{t+1|t} = 0 \quad (30)$$

en donde la expresión (26) propaga la estimación óptima del vector de estado contando con la información disponible hasta el instante anterior. Las ecuaciones (27) y (28) calculan las innovaciones, interpretables como errores de predicción a horizonte un período, y su matriz de covarianzas, respectivamente. La ganancia óptima del filtro viene dada por (29). La ecuación (30) indica que $\bar{P} = 0$ es un punto fijo de la ecuación de Riccati del filtro de Kalman, propiedad asociada a los modelos en forma *steady-state innovations* (ver Anderson y Moore, 1979) y que cumplen, por tanto, los modelos PVARMAX expresados en la forma (11)-(12).

Por su parte, el vector w_N y la matriz W_N se calculan mediante las recursiones:

$$w_t = w_{t-1} + \bar{\Phi}_{t-1}^T H_t^T B_t^{-1} \bar{z}_t \quad (31)$$

$$W_t = W_{t-1} + \bar{\Phi}_{t-1}^T H_t^T B_t^{-1} H_t \bar{\Phi}_{t-1} \quad (32)$$

$$\bar{\Phi}_t = (\bar{\Phi}_t - K_t H_t) \bar{\Phi}_{t-1} \quad (33)$$

donde w_t y W_t se inicializan a cero, $\bar{\Phi}_0 = I$ y K_t es la ganancia de un FK(0,0).

La elección adecuada de condiciones iniciales (\bar{x}_1 y P_1) para la evaluación de (25) dependerá de si el modelo es estacionario o no y de si sus variables exógenas son deterministas y/o estocásticas (ver Casals y Sotoca, 1996). En el Cuadro 1 se resumen los resultados fundamentales, haciendo referencia a la representación de coeficientes constantes (22)-(12):

Cuadro 1: Inicialización del filtro de Kalman para el cálculo de la función de verosimilitud exacta de modelos PVARMAX.

	Sin inputs	Con inputs deterministas	Con inputs estocásticos
Modelos estacionarios	$\bar{x}_1 = 0$ P_1 : solución de $P_1 = \Phi^* P_1 (\Phi^*)^T + Q^*$	$\bar{x}_1 = (I - \Phi^*)^{-1} \Gamma^* u_0$ P_1 : solución de $P_1 = \Phi^* P_1 (\Phi^*)^T + Q^*$	$\bar{x}_1 \neq 0; P_1^{-1} \neq 0^*$
Modelos no estacionarios	$\bar{x}_1 = 0; P_1^{-1} = 0$	$\bar{x}_1 \neq 0; P_1^{-1} = 0^{***}$	
Modelos parcialmente estacionarios	$\bar{x}_1 = 0; P_1^{-1} \neq 0^{***}$	$\bar{x}_1 \neq 0; P_1^{-1} \neq 0^{***}$	

* Ver Casals y Sotoca (1996).

** $\Phi^* = \prod_{j=1}^s \Phi_{s-j+1}$; $\Gamma^* = \sum_{i=1}^s (\prod_{j=1}^{s-i} \Phi_{s-j+1}) \Gamma_i$ y $Q^* = \sum_{i=1}^s [(\prod_{j=1}^{s-i} \Phi_{s-j+1}) (E_i Q_i E_i^T) (\prod_{j=1}^{s-i} \Phi_{s-j+1})^T]$

*** Ver De Jong y Chu-Chun-Lin (1994).

En los siguientes subapartados (4.1-4-3) se discute la evaluación eficiente de (25) en función de la estacionariedad o no estacionariedad del proceso y, consecuentemente, de la selección de condiciones iniciales.

4.1. Modelos PVARMAX estacionarios.

Si el modelo es estacionario y no tiene variables exógenas, la inicialización adecuada del filtro de Kalman consiste en fijar $\bar{x}_1 = 0$ y su matriz de covarianzas P_1 en la solución de la ecuación de Lyapunov correspondiente a (22), ver Cuadro 1. Por tanto, la expresión (25) se convierte en:

$$l(Z/U, \theta) = \log |P_1| + \sum_{i=1}^N \log |Q_i| + \sum_{i=1}^N \bar{z}_i^T Q_i^{-1} \bar{z}_i + \log |P_1^{-1} + W_N| - w_N^T [P_1^{-1} + W_N]^{-1} w_N \quad (34)$$

donde w_N y W_N se calculan a partir de $w_t = w_{t-1} + \bar{\Phi}_{t-1}^T H_t^T Q_t^{-1} \bar{z}_t$ y $W_t = W_{t-1} + \bar{\Phi}_{t-1}^T H_t^T Q_t^{-1} H_t \bar{\Phi}_{t-1}$, siendo $\bar{\Phi}_t = (\Phi_t - E_t H_t) \bar{\Phi}_{t-1}$ (con $\bar{\Phi}_0 = I$).

Si M es el factor Cholesky de P_1 , entonces $P_1 = MM^T$, $|P_1| = |MM^T| = |M|^2$ y $\log |P_1| + \log |P_1^{-1} + W_N| = \log |I + M^T W_N M|$. Por tanto, (34) puede escribirse como:

$$l(Z/U, \theta) = \log |I + M^T W_N M| + \sum_{i=1}^N \log |Q_i| + \sum_{i=1}^N \bar{z}_i^T Q_i^{-1} \bar{z}_i - w_N^T M [I + M^T W_N M]^{-1} M^T w_N \quad (35)$$

Si L es el factor Cholesky de $I + M^T W_N M$, tal que $(I + M^T W_N M)^{-1} = (LL^T)^{-1}$, es posible resolver el sistema de ecuaciones $L\lambda = M^T w_N$ recursivamente hacia atrás y (35) se convierte en:

$$l(Z/U, \theta) = \log |L|^2 + \sum_{i=1}^N \log |Q_i| + \sum_{i=1}^N \bar{z}_i^T Q_i^{-1} \bar{z}_i - \lambda^T \lambda \quad (36)$$

y suponiendo que N_i es el número de observaciones de la estación i -ésima y que $\sum_{i=1}^s N_i = N$ (siendo N el número total de observaciones en la muestra), el término $\sum_{i=1}^N \log |Q_i|$ de (36) puede calcularse como $\sum_{i=1}^s N_i \log |Q_i|$.

Si el modelo es estacionario y contiene variables exógenas, sean éstas deterministas o estocásticas, la función de verosimilitud sigue siendo (25), pero el filtro de Kalman debe inicializarse de forma distinta, ver Cuadro 1.

4.2. Modelos PVARMAX no estacionarios.

Si el modelo no es estacionario y no tiene variables exógenas, las ecuaciones del filtro de Kalman siguen siendo (26)-(30), pero deben inicializarse con $\bar{x}_1 = 0$ y $P_1^{-1} = 0$, lo que equivale a establecer $P_1 = kI$ siendo k una constante arbitrariamente grande. En este caso, el límite de la distancia $l(Z/U, \theta) = \log |P_1|$ cuando P_1 tiende a infinito es:

$$\sum_{i=1}^s N_i \log |Q_i| + \sum_{i=1}^N \bar{z}_i^T Q_i^{-1} \bar{z}_i + \log |W_N| - w_N^T W_N^{-1} w_N \quad (37)$$

y sólo es posible evaluar (37) para el caso no estacionario, donde w_N y W_N se obtienen, de nuevo, a partir de (31)-(33), sabiendo que $B_t = Q_t$ y $K_t = E_t \forall t = 1, 2, \dots, N$.

En modelos no estacionarios con variables exógenas, la condición inicial \bar{x}_1 es distinta de cero y su expresión dependerá del carácter determinista o estocástico de los inputs, ver Cuadro 1.

4.3. Modelos PVARMAX parcialmente no estacionarios.

Si el modelo es parcialmente no estacionario y no contiene variables exógenas, algunas raíces de $\Phi^* = \prod_{j=1}^p \Phi_{s-j+1}$ son, en módulo, mayores o iguales a la unidad y otras, menores que la unidad. En esta situación, las condiciones iniciales adecuadas son $\bar{x}_1 = 0$ y P_1^{-1} es una matriz finita y distinta de cero (ver De Jong y Chu-Chun-Lin, 1994).

La posible presencia de variables exógenas en el modelo y su carácter determinista o estocástico modifica las condiciones iniciales del filtro de Kalman y, por tanto, la evaluación de la función de verosimilitud exacta, ver Cuadro 1.

5. Esquema del algoritmo y propiedades.

5.1. Caso estacionario.

La evaluación de la función de verosimilitud exacta (36) de un proceso PVARMAX estacionario puede estructurarse en los siguientes pasos:

- 1) Calcular la factorización de Cholesky de las s matrices Q_i (es decir, calcular Q_i^{-1}), su determinante ($|Q_i| = |Q_i^{-1}|^2$) y una matriz R_i tal que $R_i Q_i R_i^T = I$ ($R_i = [Q_i^{-1}]^{-1}$).
- 2) Calcular la matriz P_1 a partir de la solución de la ecuación de Lyapunov $P_1 = \Phi^* P_1 (\Phi^*)^T + Q^*$ y su factor Cholesky M .
- 3) Calcular las secuencias \tilde{z}_t , w_t y W_t , a partir de las expresiones (26)-(30) y (31)-(33), respectivamente.
- 4) Evaluar el vector $M^T w_N$.
- 5) Calcular la matriz $I + M^T W_N M$, su factor Cholesky L y su determinante $|I + M^T W_N M| = |L|^2$.
- 6) Usar un procedimiento de *forward substitution* para resolver en λ el sistema triangular $L\lambda = M^T w_N$.
- 7) Calcular la forma cuadrática $\sum_{t=1}^N \tilde{z}_t^T Q_t^{-1} \tilde{z}_t$ y finalmente el término $\sum_{t=1}^N \tilde{z}_t^T Q_t^{-1} \tilde{z}_t - \lambda \lambda^T$. No se invierte explícitamente la matriz Q_t , sino que se usan las inversas de los s factores de Cholesky calculadas en el paso 1) (R_i , $\forall i = 1, 2, \dots, s$).

5.2. Caso no estacionario.

En el caso no estacionario, los pasos necesarios para evaluar la función de verosimilitud (37) son los siguientes:

- 1) Idéntico al caso estacionario.

- 2) No hay que darlo, ya que (37) no depende de P_1 .
- 3) Idéntico al caso estacionario.
- 4), 5) y 6) Se calculan los términos $\sum_{i=1}^s N_i \log |Q_i|$ y $\log |W_N|$ a partir de la factorización de Cholesky de W_N .
- 7) Se evalúa el término $\sum_{t=1}^N \tilde{z}_t^T Q_t^{-1} \tilde{z}_t - w_N^T W_N^{-1} w_N$ sin invertir explícitamente las matrices, ya que pueden calcularse a partir de los correspondientes factores de Cholesky.

En el caso parcialmente no estacionario debe utilizarse un esquema similar al no estacionario pero adaptando los pasos 2) a 7). En concreto, no se calcula el factor Cholesky de P_1 (denotado por M), ya que tiende a infinito. En cambio sí se evalúa el término $\log |P_1^{-1} + W_N|$ y la parte de P_1 asociada a los componentes estacionarios del sistema.

5.3. Propiedades del algoritmo.

Las principales propiedades del algoritmo definido en las secciones 5.1 y 5.2. son las siguientes:

- 1) Evita el cálculo del término $\log |P_1| + \log |P_1^{-1} + W_N|$ que aparece en (25), usando la descomposición de Cholesky de la matriz P_1 . Esto mejora sustancialmente la estabilidad numérica en la evaluación de la función de verosimilitud.
- 2) Permite detectar automáticamente situaciones de no estacionariedad. De acuerdo con la Sección 3.3, la estacionariedad del proceso puede comprobarse durante la ejecución del algoritmo inspeccionando el módulo de los autovalores de Φ^* . Esto puede hacerse eficientemente resolviendo la ecuación de Lyapunov $P_1 = \Phi^* P_1 (\Phi^*)^T + Q^*$ mediante una descomposición Schur compleja de Φ^* (ver Petkov *et al.*, 1991) que ofrece, como subproducto del proceso de cálculo, los autovalores de Φ^* .

Si P_1 tiene algún autovalor nulo, la dimensión del sistema no es mínima (ver Anderson y Moore, 1979). En este caso, conviene reducir la dimensión del vector de estado usando la descomposición en valores singulares de P_1 .

- 3) Permite detectar automáticamente situaciones de no invertibilidad. De acuerdo con la Sección 3.3, se comprueba si la matriz $\bar{\Phi}_t = (\Phi_t - E_t H_t) \bar{\Phi}_{t-1}$ converge a cero cuando $t \rightarrow N$, siendo N el tamaño de la muestra.
- 4) Si a_t y x_t son variables aleatorias normales independientes, las innovaciones generadas por el filtro (26)-(30) se distribuyen como una normal con esperanza nula y, cuando $t \rightarrow N$, convergen en media cuadrática a las perturbaciones a_t (Shea, 1989, pág. 162). Consecuentemente, estas innovaciones son adecuadas como instrumentos para la diagnosis del modelo.

6. Resultados con datos simulados.

En este apartado se presentan los resultados de un conjunto de ejercicios de simulación que ilustran el buen funcionamiento y la eficiencia del procedimiento propuesto.

El primer ejercicio utiliza la misma estructura $\text{PARMA}_2(1,1)$ que en los trabajos de Vecchia (1985) y Li y Hui (1988): $(1 + f_{t,1}B)z_t = (1 + \theta_{t,1}B)a_t$, con $f_{t,1} = f_{t+s,1}$, $\theta_{t,1} = \theta_{t+s,1}$ y $\sigma_t^2 = \sigma_{t+s}^2$, $s=2$ y $t = 1, 2, \dots, N$. Por tanto, para la primera estación se tiene:

$$(1 + f_{1,1}B)z_t = (1 + \theta_{1,1}B)a_{1,t}, \quad a_{1,t} \sim \text{IID}(0, q_{1,1}), \quad t = 1, 3, 5, \dots \quad (38)$$

mientras que el modelo de la segunda estación es:

$$(1 + f_{2,1}B)z_t = (1 + \theta_{2,1}B)a_{2,t}, \quad a_{2,t} \sim \text{IID}(0, q_{2,1}), \quad t = 2, 4, 6, \dots \quad (39)$$

A partir de esta formulación, se simulan muestras de 100 y 200 observaciones, con 150 repeticiones para cada tamaño muestral y considerando distintos valores teóricos de los parámetros. En las tablas 1.a-1.c se ofrecen, para cada tamaño muestral, la media de las estimaciones, la raíz cuadrada del error cuadrático medio (RMSE), la desviación típica muestral (DTM) y la desviación típica exacta (DTE), calculada sustituyendo los valores teóricos de los parámetros en la matriz de información exacta (ver Apéndice). Comparando esta cifra con la desviación típica muestral, se obtiene una idea de la eficiencia de la estimación y se puede comprobar si las desviaciones típicas muestrales convergen a las exactas, calculadas a partir de la cota de Cramer-Rao.

Tanto el RMSE como el sesgo de la estimación por máxima verosimilitud exacta son menores, y en algunos casos mucho menores, que los ofrecidos por Vecchia (1985) y prácticamente iguales que los de Li y Hui (1988). También cabe destacar cómo la eficiencia del estimador, medida por el RMSE, mejora al aumentar el tamaño muestral.

Para probar el comportamiento del algoritmo cuando las estructuras estocásticas propias de cada estación son diferentes, se ha simulado el proceso: $(1 + f_{t,1}B + f_{t,2}B^2)z_t = (1 + \theta_{t,1}B)a_t$, con $s=2$ y $t = 1, 2, \dots, N$, de forma que para la primera estación se tiene:

$$(1 + f_{1,1}B + f_{1,2}B^2)z_t = (1 + \theta_{1,1}B)a_{1,t}, \quad a_{1,t} \sim \text{IID}(0, q_{1,1}), \quad t = 1, 3, 5, \dots \quad (40)$$

con $f_{1,2} = 0$ mientras que el modelo de la segunda estación es:

$$(1 + f_{2,1}B + f_{2,2}B^2)z_t = (1 + \theta_{2,1}B)a_{2,t}, \quad a_{2,t} \sim \text{IID}(0, q_{2,1}), \quad t = 2, 4, 6, \dots \quad (41)$$

con $\theta_{2,1} = 0$. En las Tablas 2.a-2.c se muestran los resultados correspondientes a distintos valores teóricos de los parámetros, para los mismos tamaños muestrales y el mismo número de repeticiones que el ejercicio anterior. Las conclusiones también son análogas.

Por último, en la Tabla 3 se muestran los resultados de la estimación de un modelo bivalente definido, siguiendo (1)-(4) como: $(I + F_{s,1}B)z_t = (I + \Theta_{s,1}B)a_t$, con $s=2$ y $t = 1, 2, \dots, N$, de forma que para la primera estación el modelo es:

$$\begin{bmatrix} 1 + f_{11}^1 B & f_{12}^1 B \\ 0 & 1 + f_{22}^1 B \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 + \theta_{11}^1 B & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}, \quad E \left\{ \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \begin{bmatrix} a_{1,t} & a_{2,t} \end{bmatrix} \right\} = \begin{bmatrix} q_{11}^1 & 0 \\ 0 & q_{22}^1 \end{bmatrix}, \quad t = 1, 3, 5, \dots \quad (42)$$

mientras que para la segunda estación se tiene:

$$\begin{bmatrix} 1 + f_{11}^2 B & 0 \\ f_{21}^2 B & 1 + f_{22}^2 B \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 + \theta_{22}^2 B \end{bmatrix} \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}, \quad E \left\{ \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \begin{bmatrix} a_{1,t} & a_{2,t} \end{bmatrix} \right\} = \begin{bmatrix} q_{11}^2 & 0 \\ 0 & q_{22}^2 \end{bmatrix}, \quad t = 2, 4, 6, \dots \quad (43)$$

Los resultados del ejercicio se resumen en la Tabla 3, en donde nuevamente se observa que las estimaciones son adecuadas, tanto en términos de sesgo como de RMSE.

Tabla 1.a: Resultados de la estimación por máxima verosimilitud exacta del modelo (38)-(39).

	$f_{1,1} = -.5$	$\theta_{1,1} = -.9$	$q_{1,1} = 16.0$	$f_{2,1} = -1.4$	$\theta_{2,1} = .7$	$q_{2,1} = 64.0$
Media						
N=100						
RMSE†	.497	.898	14.978	-1.391	.720	62.113
DTM†	.083	.112	3.014	.148	.174	12.294
DTE†	.083	.112	2.835	.147	.172	12.148
	.085	.106	3.218	.135	.147	12.810
N=200						
RMSE†	.491	.892	15.743	-1.409	.716	64.335
DTM†	.064	.077	2.157	.099	.115	9.232
DTE†	.063	.076	2.141	.099	.114	9.226
	.061	.076	2.269	.096	.103	9.054

† RMSE: raíz cuadrada del error cuadrático medio; DTM: desviación típica muestral; DTE: desviación típica exacta.

Tabla 1.b: Resultados de la estimación por máxima verosimilitud exacta del modelo (38)-(39).

		$f_{1,1} = -.8$	$\theta_{1,1} = .5$	$q_{1,1} = 4.0$	$f_{2,1} = -.5$	$\theta_{2,1} = -.8$	$q_{2,1} = 1.0$
N=100	Media	-.790	.528	3.883	-.534	-.837	.935
	RMSE†	.225	.273	.847	.222	.233	.229
	DTM†	.225	.272	.838	.219	.230	.220
	DTE†	.209	.231	.803	.175	.190	.200
N=200	Media	-.806	.497	3.877	-.510	-.810	.995
	RMSE†	.159	.148	.604	.134	.142	.146
	DTM†	.159	.148	.591	.133	.142	.146
	DTE†	.147	.162	.567	.124	.135	.142

† RMSE: raíz cuadrada del error cuadrático medio; DTM: desviación típica muestral; DTE: desviación típica exacta.

Tabla 1.c: Resultados de la estimación por máxima verosimilitud exacta del modelo (38)-(39).

		$f_{1,1} = -.8$	$\theta_{1,1} = -.3$	$q_{1,1} = 1.0$	$f_{2,1} = -.6$	$\theta_{2,1} = -.5$	$q_{2,1} = 1.0$
N=100	Media	-.754	-.239	.960	-.649	-.548	.968
	RMSE†	.200	.217	.213	.217	.288	.218
	DTM†	.195	.208	.209	.211	.284	.216
	DTE†	.383	.394	.201	.275	.304	.200
N=200	Media	-.753	-.257	.961	-.636	-.538	.990
	RMSE†	.147	.144	.144	.168	.193	.136
	DTM†	.139	.138	.138	.164	.190	.136
	DTE†	.269	.277	.142	.194	.214	.141

† RMSE: raíz cuadrada del error cuadrático medio; DTM: desviación típica muestral; DTE: desviación típica exacta.

Tabla 2.a: Resultados de la estimación por máxima verosimilitud exacta del modelo (40)-(41).

		$f_{1,1} = -.5$	$\theta_{1,1} = -.9$	$q_{1,1} = 16.0$	$f_{1,2} = -.2$	$f_{2,2} = -.7$	$q_{2,1} = 64.0$
$N=100$	Media	-.555	-.958	14.903	-.183	-.663	62.278
	RMSE†	.158	.174	3.408	.094	.096	12.509
	DTM†	.168	.184	3.580	.095	.103	13.036
	DTE†	.138	.154	3.256	.088	.093	12.714
$N=200$	Media	-.526	-.932	15.581	-.184	-.685	64.275
	RMSE†	.125	.136	2.299	.064	.075	9.714
	DTM†	.127	.140	2.336	.066	.077	9.718
	DTE†	.098	.110	2.282	.062	.067	9.019

† RMSE: raíz cuadrada del error cuadrático medio; DTM: desviación típica muestral; DTE: desviación típica exacta.

Tabla 2.b: Resultados de la estimación por máxima verosimilitud exacta del modelo (40)-(41).

		$f_{1,1} = -.8$	$\theta_{1,1} = .5$	$q_{1,1} = 4.0$	$f_{2,1} = -.7$	$f_{2,2} = .5$	$q_{2,1} = 1.0$
$N=100$	Media	-.810	.486	3.680	-.698	.501	.966
	RMSE†	.205	.340	.741	.075	.106	.204
	DTM†	.206	.341	.807	.075	.106	.206
	DTE†	.201	.351	.803	.070	.102	.201
$N=200$	Media	-.781	.512	3.858	-.701	.492	.985
	RMSE†	.141	.261	.552	.051	.077	.140
	DTM†	.142	.261	.570	.051	.078	.141
	DTE†	.142	.247	.567	.049	.072	.142

† RMSE: raíz cuadrada del error cuadrático medio; DTM: desviación típica muestral; DTE: desviación típica exacta.

Tabla 2.c: Resultados de la estimación por máxima verosimilitud exacta del modelo (40)-(41).

		$f_{1,1} = -.6$	$\theta_{1,1} = -.5$	$q_{1,1} = 1.0$	$f_{1,2} = -.2$	$f_{2,2} = -.7$	$q_{2,1} = 1.0$
N=100	Media	-.626	-.538	.964	-.180	-.675	.976
	RMSE†	.166	.241	.202	.121	.105	.173
	DTM†	.168	.244	.205	.122	.108	.174
	DTE†	.122	.189	.201	.119	.096	.200
N=200	Media	-.609	-.512	.973	-.205	-.680	1.001
	RMSE†	.095	.144	.157	.080	.078	.150
	DTM†	.095	.144	.157	.080	.080	.150
	DTE†	.086	.133	.142	.084	.068	.141

† RMSE: raíz cuadrada del error cuadrático medio; DTM: desviación típica muestral; DTE: desviación típica exacta.

Tabla 3: Resultados de la estimación por máxima verosimilitud exacta del modelo (42)-(43).

		$f_{11}^1 = -.8$	$f_{12}^1 = -.3$	$f_{22}^1 = -.5$	$\theta_{11}^1 = -.4$	$q_{11}^1 = 1.0$	$q_{22}^1 = .5$	$f_{11}^2 = -.7$	$f_{21}^2 = -.3$	$f_{22}^2 = -.5$	$\theta_{22}^2 = -.8$	$q_{11}^2 = .5$	$q_{22}^2 = 1.0$
N= 100	Media	-.787	-.295	-.498	-.371	.923	.490	-.688	-.291	-.483	-.791	.483	.953
	RMSE†	.177	.143	.087	.248	.217	.099	.072	.130	.281	.359	.092	.206
	DTM†	.176	.143	.087	.246	.203	.098	.071	.130	.280	.359	.091	.200
	DTE†	.162	.135	.083	.260	.201	.100	.067	.112	.305	.366	.100	.200
N= 200	Media	-.805	-.290	-.499	-.415	.975	.499	-.693	-.305	-.494	-.825	.503	.968
	RMSE†	.113	.098	.061	.176	.141	.071	.049	.076	.234	.282	.073	.140
	DTM†	.113	.097	.061	.176	.139	.017	.048	.076	.234	.281	.073	.137
	DTE†	.114	.095	.058	.183	.142	.071	.048	.079	.215	.258	.071	.141
N= 500	Media	-.795	-.294	-.499	-.405	.987	.496	-.700	-.306	-.497	-.797	.496	.991
	RMSE†	.073	.062	.043	.118	.071	.041	.029	.052	.141	.164	.046	.093
	DTM†	.073	.062	.043	.118	.070	.041	.029	.052	.141	.164	.045	.093
	DTE†	.072	.060	.037	.115	.090	.045	.030	.050	.136	.163	.045	.089

† RMSE: raíz cuadrada del error cuadrático medio; DTM: desviación típica muestral; DTE: desviación típica exacta.

7. Conclusiones.

La literatura econométrica ha mostrado un interés creciente por los modelos periódicos, debido a su utilidad para predecir series temporales estacionales. Sin embargo, su estimación es compleja, por el elevado número de parámetros que los caracterizan y costosa en términos de tiempo de cálculo. En este trabajo se propone un procedimiento computacionalmente eficiente para estimar modelos periódicos por máxima verosimilitud exacta.

Además del coste computacional, los métodos estándar adolecen de falta de flexibilidad ya que: a) tratan de forma diferenciada la estimación de modelos univariantes y multivariantes, b) sólo contemplan el caso estacionario, c) no consideran la posible existencia de variables exógenas en el modelo, d) suponen el mismo número de observaciones por estación y e) cuando no todas las estaciones presentan la misma estructura dinámica incurrir en ineficiencia, ya que necesitan aumentar artificialmente la dinámica del sistema.

Las contribuciones fundamentales de este trabajo consisten en la definición de: a) un proceso estocástico periódico generalizado, b) un procedimiento de conversión del mismo a una forma equivalente en espacio de los estados de dimensión mínima y c) un algoritmo eficiente para el cálculo de su función de verosimilitud exacta. Estas aportaciones configuran un proceso de modelización y cálculo que, frente a los existentes en la literatura, presenta ventajas en términos de flexibilidad de la representación, eficiencia computacional y capacidad para aplicar resultados conocidos en el contexto de modelos en espacio de los estados.

En cuanto al primer aspecto (flexibilidad de la representación) el trabajo parte de la definición de una familia de procesos estocásticos VARMAX periódicos (PVARMAX) más general que las ya existentes. El proceso PVARMAX puede entenderse como un proceso VARMAX en el que los parámetros cambian de forma periódica. Esta formulación es general incluyendo tanto modelos periódicos univariantes como multivariantes, cuando la literatura trata de forma diferenciada ambas situaciones. Además, el uso de una representación en EE equivalente a la formulación PVARMAX permite que la muestra comience y termine en cualquier estación, no siendo necesario que incluya un número entero de ciclos estacionales.

En cuanto al segundo aspecto (eficiencia computacional), ésta se consigue combinando a) una representación del proceso PVARMAX como un modelo en espacio de los estados de

dimensión mínima en forma *steady-state innovations* y b) un algoritmo computacionalmente eficiente para calcular la función de verosimilitud exacta, que aprovecha las propiedades de convergencia del filtro de Kalman cuando se aplica a modelos en forma *steady-state innovations*. Así se consigue que el número de operaciones usado en la estimación de un $\text{PARMAX}_s(p_i, q_i, g_i)$, donde $p_i = p$, $q_i = q$, $g_i = g$, $\forall i = 1, 2, \dots, s$, s es el número de estaciones consideradas y g el número de variables exógenas, no difiera sustancialmente del correspondiente a la estimación de una estructura $\text{ARMAX}_s(p, q, g)$ univariante.

En cuanto al tercer y último aspecto (capacidad para aplicar resultados conocidos) el uso de una representación en espacio de los estados facilita la aplicación de resultados existentes para a) tratar modelos estacionarios o no estacionarios, con variables explicativas estocásticas y/o deterministas (Casals y Sotoca, 1997), b) detectar automáticamente posibles situaciones de no estacionariedad y no invertibilidad, de forma similar al criterio de Mauricio (1995) para detección de no invertibilidad y c) calcular la matriz de información analítica y, por tanto, los segundos momentos exactos de las estimaciones máximo-verosímiles (Terceiro, 1990 y Apéndice).

Los resultados obtenidos con datos simulados muestran el buen funcionamiento del algoritmo en la estimación de modelos periódicos univariantes y multivariantes.

Por último, cabe destacar que todos estos resultados pueden aplicarse inmediatamente a la estimación de modelos VARMAX con cualquier pauta de variación paramétrica sistemática o de regímenes cambiantes, siendo la formulación periódica un caso particular de este planteamiento general.

Referencias

- Adams, G.J. y Goodwin, G.C. (1995). Parameter Estimation for Periodic ARMA Models. *Journal of Time Series Analysis*, 16, 2, 127-145.
- Anderson, B.D.O. y Moore, J.B. (1979). *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall.
- Boshnakov, G.N. (1996). Recursive Computation of the Parameters of Periodic Autoregressive Moving-Average Processes. *Journal of Time Series Analysis*, 17, 4, 333-349.
- Boswijk, P. y Franses, P.H. (1995). Unit roots in Periodic Autorregressions. *Journal of Time Series Analysis*, 17, 3, 221-245.
- Casals, J. y Sotoca, S. (1998). Exact Initial Conditions for Maximum Likelihood Estimation of State Space Models with Stochastic Inputs. *Economics Letters*, forthcoming.
- De Jong, P. (1988). The Likelihood for a State Space Model. *Biometrika* 75, 1, 165-169.
- De Jong, P. y Chu-Chun-Lin, S. (1994). Stationary and Non-Stationary State Space Models. *Journal of Time Series Analysis*, 15, 2, 151-166.
- Franses, P.H. y Paap, R. (1994). Model Selection in Periodic Autorregressions. *Oxford Bulletin of Economics and Statistics*, 56, 4, 421-439.
- Gardner, G., Harvey, A.C. y Phillips, G.D.A. (1980). Algorithm 154. An Algorithm for Exact Maximum-Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering. *Applied Statistics*, 29, 311-317.
- Hannan, E.J. y Kavalieris, L. (1984). Multivariate Linear Time Series Models. *Advances in Applied Probability*, 16, 492-561.
- Hannan, E.J. y Rissanen, J. (1982). Recursive Estimation of Mixed Autoregressive-Moving Average Model. *Biometrika*, 69, 81-94.
- Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin. (second edition).
- Li, W.K. y Hui, Y.V. (1988). An Algorithm for the Exact Likelihood of Periodic Autoregressive Moving Average Models. *Commun. Statist.-Simula.*, 17, 4, 1483-1494.
- Mauricio, A. (1995). Exact Maximum Likelihood Estimation of Stationary Vector ARMA Models. *Journal of the American Statistical Association*, 90, 429, 282-291.
- McLeod, A.L. (1975). Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Time Series. *Applied Statistics*, 24, 255-256.
- Noakes, D.J., McLeod, A.L. y Hipel, K. (1985). Forecasting Monthly Riverflow Time Series. *International Journal of Forecasting*, 1, 179-190.
- Osborn, D.R. (1991). The Implications of Periodically Varying Coefficients for Seasonal Time-Series Processes. *Journal of Econometrics*, 48, 373-384.
- Pagano, M. (1978). On Periodic and Multiple Autoregression. *Annals of Statistics*, 6, 1310-1317.
- Petkov, P.Hr., Christov, N.D. y Konstantinov, M.M. (1991). *Computational Methods for Linear Control Systems*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Salas, J.D., Boes, D.C. y Smith, R.A. (1982). Estimation of ARMA Models with Seasonal Parameters. *Water Resources Research*, 18, 1006-1010.
- Shea, B.L. (1989). Algorithm AS 242. The Exact Likelihood of a Vector Autoregressive-Moving Average Model. *Applied Statistics*, 38, 161-204.
- Terceiro, J. (1990). *Estimation of Dynamic Econometric Models with Errors in Variables*. Springer-Verlag, Heidelberg.
- Vecchia, A.V. (1985). Maximum Likelihood Estimation for Periodic Autoregressive Moving Average Models. *Technometrics*, 27, 4, 375-384.

Apéndice: Cálculo eficiente de la matriz de Información.

Una forma equivalente de escribir la función de verosimilitud (25), sin tener en cuenta constantes, es:

$$l(Z/U, \theta) = \sum_{t=1}^N \log |B_t| + \sum_{t=1}^N \tilde{z}_t^T B_t^{-1} \tilde{z}_t \quad (1.1)$$

donde \tilde{z}_t y B_t son las innovaciones y su matriz de covarianzas resultantes del siguiente filtro de Kalman inicializado adecuadamente:

$$\tilde{z}_t = z_t - H_t \hat{x}_{t|t-1} - D_t u_t \quad (1.2)$$

$$\hat{x}_{t+1|t} = \Phi_t \hat{x}_{t|t-1} + \Gamma_t u_t + K_t \tilde{z}_t \quad (1.3)$$

$$K_t = (\Phi_t P_{t|t-1} H_t^T + E_t Q_t) B_t^{-1} \quad (1.4)$$

$$P_{t+1|t} = \Phi_t P_{t|t-1} \Phi_t^T + E_t Q_t E_t^T - K_t B_t K_t^T \quad (1.5)$$

$$B_t = H_t P_{t|t-1} H_t^T + Q_t \quad (1.6)$$

A partir de (1.1), es fácil ver que la expresión analítica del gradiente de la función de verosimilitud es:

$$\frac{\partial l(\theta)}{\partial \theta_i} = \sum_{t=1}^N \text{tr} \left[B_t^{-1} \frac{\partial B_t}{\partial \theta_i} \right] + 2 \sum_{t=1}^N \tilde{z}_t^T B_t^{-1} \frac{\partial \tilde{z}_t}{\partial \theta_i} - \sum_{t=1}^N \tilde{z}_t^T B_t^{-1} \frac{\partial B_t}{\partial \theta_i} B_t^{-1} \tilde{z}_t \quad (1.7)$$

donde las derivadas del vector de innovaciones y de su matriz de covarianzas se obtienen teniendo en cuenta (1.2)-(1.6):

$$\frac{\partial \tilde{z}_t}{\partial \theta_i} = -H_t \frac{\partial \hat{x}_{t|t-1}}{\partial \theta_i} - \frac{\partial D_t}{\partial \theta_i} u_t \quad (1.8)$$

$$\frac{\partial \hat{x}_{t+1|t}}{\partial \theta_i} = \frac{\partial \Phi_t}{\partial \theta_i} \hat{x}_{t|t-1} + \bar{\Phi}_t \frac{\partial \hat{x}_{t|t-1}}{\partial \theta_i} + \frac{\partial \Gamma_t}{\partial \theta_i} u_t + \frac{\partial K_t}{\partial \theta_i} \tilde{z}_t \quad (1.9)$$

siendo $\bar{\Phi}_t = \Phi_t - K_t H_t$ y $\bar{\Gamma}_t = \Gamma_t - K_t D_t$. La propagación de (1.9) requiere calcular la derivada de la ganancia del filtro de Kalman. A partir de (1.4) se tiene que:

$$\frac{\partial K_t}{\partial \theta_i} = \left[\frac{\partial \Phi_t}{\partial \theta_i} P_{t|t-1} H_t^T + \Phi_t \frac{\partial P_{t|t-1}}{\partial \theta_i} H_t^T + \frac{\partial E_t}{\partial \theta_i} Q_t + E_t \frac{\partial Q_t}{\partial \theta_i} \right] B_t^{-1} - K_t \frac{\partial B_t}{\partial \theta_i} B_t^{-1} \quad (1.10)$$

$$\frac{\partial B_t}{\partial \theta_i} = H_t \frac{\partial P_{t|t-1}}{\partial \theta_i} H_t^T + \frac{\partial Q_t}{\partial \theta_i} \quad (1.11)$$

$$\frac{\partial P_{t+1|t}}{\partial \theta_i} = \bar{\Phi}_t \frac{\partial P_{t|t-1}}{\partial \theta_i} \bar{\Phi}_t^T + A_{it} + A_{it}^T \quad (1.12)$$

siendo

$$A_{it} = \frac{\partial \Phi_t}{\partial \theta_i} P_{t|t-1} \Phi_t^T + \frac{\partial E_t}{\partial \theta_i} Q_t K_t^T + E_t \frac{\partial Q_t}{\partial \theta_i} K_t^T + \frac{1}{2} \frac{\partial (E_t Q_t E_t^T)}{\partial \theta_i} + \frac{1}{2} K_t \frac{\partial Q_t}{\partial \theta_i} K_t^T \quad (1.13)$$

Un instrumento necesario para validar el modelo estimado es la matriz de covarianzas de las estimaciones. Ésta puede obtenerse a partir de la matriz de información del modelo, ya que su inversa proporciona una cota inferior a la matriz de covarianzas de los estimadores de los parámetros. El término general de la matriz de información, Terceiro (1990), es:

$$M(\theta)_{ij} = \sum_{t=1}^N \frac{1}{2} \text{tr} \left[B_t^{-1} \frac{\partial B_t}{\partial \theta_i} B_t^{-1} \frac{\partial B_t}{\partial \theta_j} \right] + \sum_{t=1}^N \text{tr} \left[B_t^{-1} E \left[\frac{\partial \tilde{z}_t}{\partial \theta_i} \frac{\partial \tilde{z}_t^T}{\partial \theta_j} \right] \right] \quad (1.14)$$

donde se ha tenido en cuenta que el proceso de innovaciones tiene esperanza nula y que la derivada de este proceso no depende de las innovaciones. En (1.14) el único término desconocido es la esperanza del producto de las derivadas del vector de innovaciones. A menudo se aproxima la esperanza de ese producto por el producto de las derivadas. Sin embargo, para calcularlo de forma exacta, puede tenerse en cuenta que:

$$E \left[\frac{\partial \tilde{z}_t}{\partial \theta_i} \frac{\partial \tilde{z}_t^T}{\partial \theta_j} \right] = B_t^{ij} + \bar{z}_t \bar{z}_t^T \quad (1.15)$$

donde \bar{z}_t es la media de $\partial \tilde{z}_t / \partial \theta_i$ y B_t^{ij} es la matriz de covarianzas de $\partial \tilde{z}_t / \partial \theta_i$ y $\partial \tilde{z}_t / \partial \theta_j$. El problema queda reducido, por tanto, a calcular los dos sumandos del lado derecho de (1.15).

Para ello, se define el sistema ampliado:

$$x_{t+1}^c = \Phi_t^c x_t^c + \Gamma_t^c u_t + K_t^c \tilde{z}_t \quad (1.16)$$

$$z_t^c = H_t^c x_t^c + D_t^c u_t \quad (1.17)$$

donde:

$$x_t^c = \begin{bmatrix} \hat{x}_{t-1} \\ \frac{\partial \hat{x}_{t-1}}{\partial \theta} \\ \frac{\partial \hat{x}_{t-1}}{\partial \theta} \end{bmatrix}, \Phi_t^c = \begin{bmatrix} \Phi_t & 0 & 0 \\ \frac{\partial \Phi_t}{\partial \theta} & \bar{\Phi}_t & 0 \\ \frac{\partial \Phi_t}{\partial \theta} & 0 & \bar{\Phi}_t \end{bmatrix}, \Gamma_t^c = \begin{bmatrix} \Gamma_t \\ \frac{\partial \Gamma_t}{\partial \theta} - K_t \frac{\partial D_t}{\partial \theta} \\ \frac{\partial \Gamma_t}{\partial \theta} - K_t \frac{\partial D_t}{\partial \theta} \end{bmatrix}, K_t^c = \begin{bmatrix} K_t \\ \frac{\partial K_t}{\partial \theta} \\ \frac{\partial K_t}{\partial \theta} \end{bmatrix} \quad (1.18)$$

$$z_t^c = \begin{bmatrix} \frac{\partial z_t}{\partial \theta} \\ \frac{\partial z_t}{\partial \theta} \end{bmatrix}, H_t^c = \begin{bmatrix} 0 & -H_t & 0 \\ 0 & 0 & -H_t \end{bmatrix}, D_t^c = \begin{bmatrix} \frac{\partial D_t}{\partial \theta} \\ \frac{\partial D_t}{\partial \theta} \end{bmatrix} \quad (1.19)$$

Si se denota por $\bar{x}_t^c = E(x_t^c)$, $\bar{z}_t^c = E(z_t^c)$, $P_t^c = E[(x_t^c - \bar{x}_t^c)(x_t^c - \bar{x}_t^c)^T]$ y $B_t^c = E[(z_t^c - \bar{z}_t^c)(z_t^c - \bar{z}_t^c)^T]$, las ecuaciones que permiten calcular estos momentos son:

$$\bar{x}_{t+1}^c = \Phi_t^c \bar{x}_t^c + \Gamma_t^c u_t \quad (1.20)$$

$$\bar{z}_t^c = H_t^c \bar{x}_t^c + D_t^c u_t \quad (1.21)$$

$$P_{t+1}^c = \Phi_t^c P_t^c (\Phi_t^c)^T + K_t^c B_t (K_t^c)^T \quad (1.22)$$

$$B_t^c = H_t^c P_t^c (H_t^c)^T \quad (1.23)$$

La propagación de (1.20) y (1.21) permite calcular en cada instante los valores de \bar{z}_t y \bar{z}_t^c , y las ecuaciones (1.22) y (1.23) proporcionan el valor de la matriz B_t^c , ya que la estructura por bloques de B_t^c es:

$$B_t^c = \begin{bmatrix} B_t^u & B_t^y \\ B_t^u & B_t^y \end{bmatrix}$$

Las condiciones iniciales del sistema dado por (1.20)-(1.23) son $\bar{x}_1^c = [\bar{x}_1^T \ 0 \ 0]^T$ y $P_1^c = \text{diag}[P_1, 0, 0]$ donde el cálculo correcto de \bar{x}_1 y P_1 se ha discutido en la Sección 4.

La matriz de información puede calcularse eficientemente descomponiendo el sistema (1.20)-(1.23) en un subsistema de dimensión menor, que permita calcular las variables \bar{z}_u , \bar{z}_y

y B_t^{ij} . Así, descomponiendo la matriz P_t^c en 3×3 bloques de dimensión n_i , de forma que P_t^{ij} se corresponde con el bloque (i, j) , se tiene que:

$$B_t^{ij} = H_i P_t^{c23} H_j^T \quad (1.24)$$

y dada la estructura de Φ_t^c , el término P_t^{c23} de (1.24) se calcula usando las expresiones:

$$P_{t+1}^{c23} = \frac{\partial \bar{\Phi}_t}{\partial \theta} P_t^{c11} \frac{\partial \bar{\Phi}_t^T}{\partial \theta} + \bar{\Phi}_t P_t^{c21} \frac{\partial \bar{\Phi}_t^T}{\partial \theta} + \frac{\partial \bar{\Phi}_t}{\partial \theta} P_t^{c13} \bar{\Phi}_t^T + \bar{\Phi}_t P_t^{c23} \bar{\Phi}_t^T + \frac{\partial K_t}{\partial \theta} B_t \frac{\partial K_t^T}{\partial \theta} \quad (1.25)$$

$$P_{t+1}^{c11} = \Phi_t P_t^{c11} \Phi_t^T + K_t B_t K_t^T \quad (1.26)$$

$$P_{t+1}^{c21} = \frac{\partial \bar{\Phi}_t}{\partial \theta} P_t^{c11} \Phi_t^T + \bar{\Phi}_t P_t^{c21} \Phi_t^T + \frac{\partial K_t}{\partial \theta} B_t K_t^T \quad (1.27)$$

$$P_{t+1}^{c13} = \Phi_t P_t^{c11} \frac{\partial \bar{\Phi}_t^T}{\partial \theta} + \bar{\Phi}_t P_t^{c13} \bar{\Phi}_t^T + K_t B_t \frac{\partial K_t^T}{\partial \theta} \quad (1.28)$$

donde puede observarse que no aparecen los bloques (2,2) y (3,3) de P_t^c . Por tanto, la carga computacional que supone la propagación de (1.25)-(1.28) es equivalente a la de un sistema de dimensión $2n_i$ no simétrico. En cuanto a la propagación de (1.20) y (1.21) para obtener \bar{z}_u y \bar{z}_y , se calcula la secuencia de medias una sola vez para cada coeficiente estimado de la siguiente forma:

$$\bar{x}_{t+1} = \Phi_t \bar{x}_t + \Gamma_t u_t \quad (1.29)$$

$$\bar{x}_{t+1,u} = \bar{\Phi}_t \bar{x}_{t,u} + \frac{\partial \bar{\Phi}_t}{\partial \theta} \bar{x}_t + \frac{\partial \Gamma_t}{\partial \theta} u_t - K_t \frac{\partial D_t}{\partial \theta} u_t \quad (1.30)$$

$$\bar{z}_{t,u} = -H_t \bar{x}_{t,u} - \frac{\partial D_t}{\partial \theta} u_t \quad (1.31)$$

donde $\bar{x}_{t+1} = E(\hat{x}_{t+1,u})$ y $\bar{x}_{t+1,u} = E(\partial \hat{x}_{t+1,u} / \partial \theta)$. La descomposición efectuada en (1.15) mejora la eficiencia en los cálculos. La obtención de B_t^{ij} supone una menor carga computacional que la correspondiente a la obtención directa de la esperanza del producto de las derivadas de las innovaciones, dado que la primera de estas expresiones es más simple al cancelarse las variables exógenas por tratarse de un momento centrado en el valor medio. Dado que el número de propagaciones es igual a $h(h-1)/2$ donde h es la dimensión del vector θ ,

la carga que pueda suponer la propagación h veces de (1.29)-(1.31) queda sobradamente compensada.

Si no existen variables exógenas en el modelo, puede suponerse que la esperanza del producto de las derivadas del proceso de innovaciones coincide con la matriz B_t^y , tomando como condiciones iniciales $P_1^c = \text{diag}\{P_1 + \bar{x}_1 \bar{x}_1^T, 0, 0\}$. Por tanto, en este caso, no sería necesario propagar el sistema (1.29)-(1.31).